

# Just the Facts: **Technical Differences Between Talend Data Integration and Talend Open Studio**

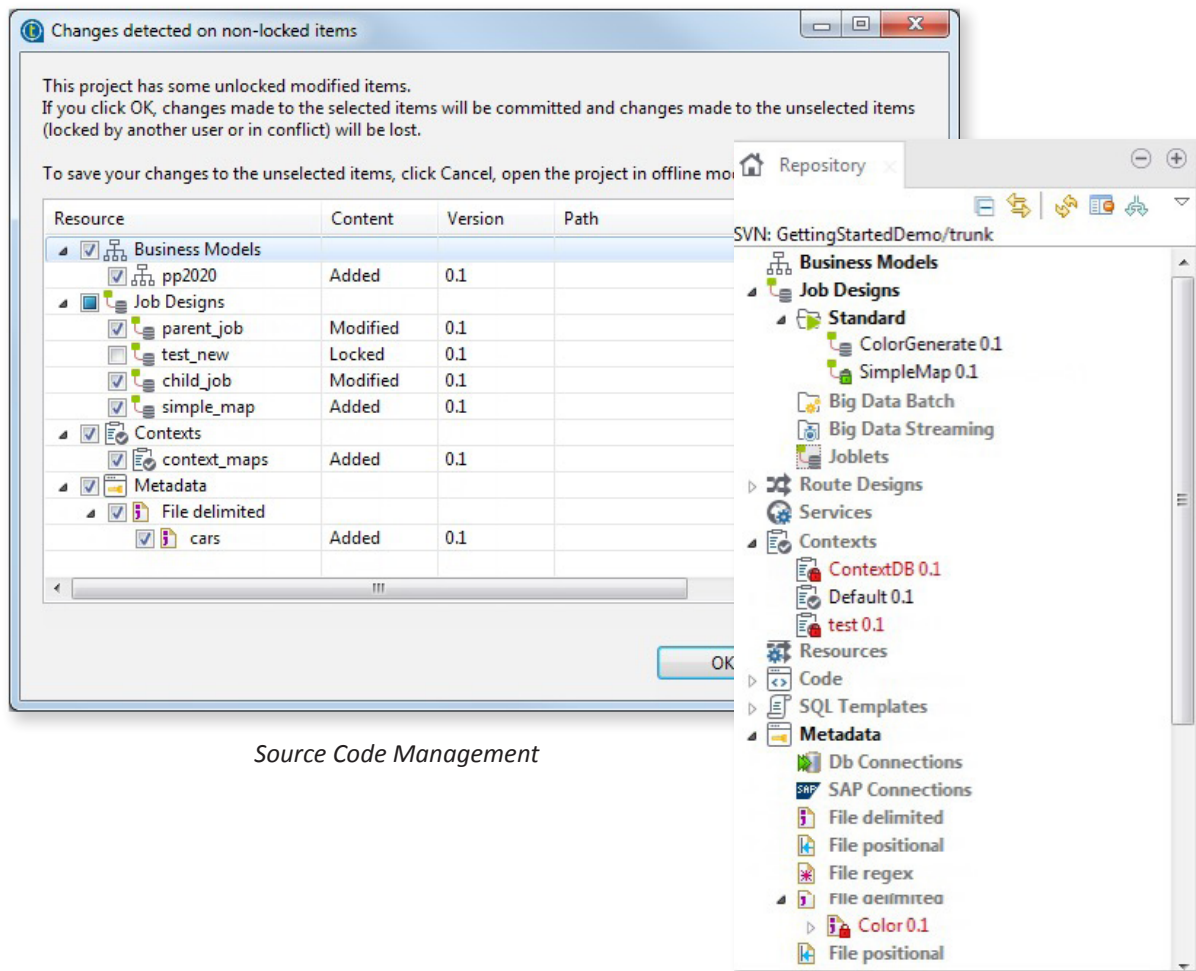
September 2017



# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

## 1. Project and Source Control Management

No more developing in silos. With Talend Data Integration, you start by defining projects, users and roles in the Talend Administration Center. Next with source code management, your team does more by storing and sharing business models, data integration jobs, metadata, and documentation in a centralized repository (GIT, SVN). When you save your job, your colleagues immediately see your work and changes, and can check in/out the job for further updates.

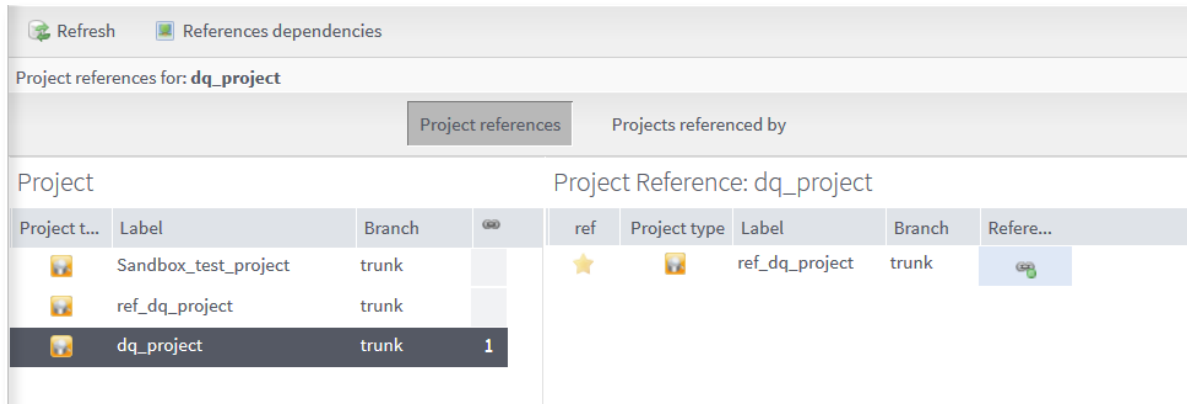


Source Code Management

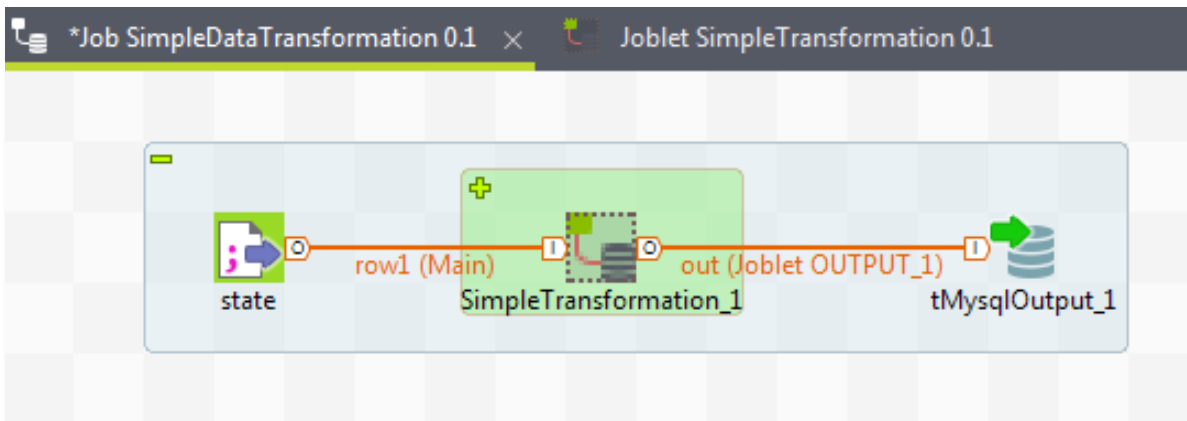
Shared Repository View Using SVN

**Reference Projects:** To increase reuse and avoid duplication of items across jobs, reference projects of common objects and shared code can be created, which can then be included or linked to other projects. When the master changes, all slave projects are updated.

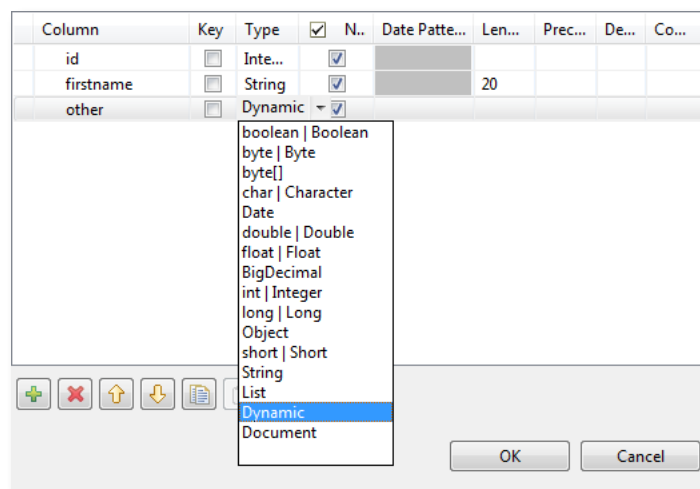
# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio



**Joblets:** Simplify complex jobs by creating joblets which factorize job parts for reuse, like simple or complex transformations.

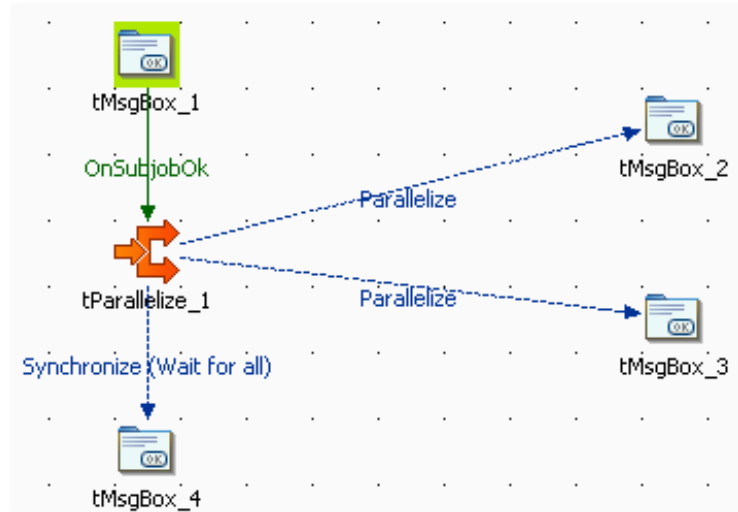


**Dynamic Schemas:** When the schema of the target is not known at design time, or when the target schema may vary depending on the specific instance of the database, dynamic schemas can be used to map source data to these targets “on-the-fly”. This reduces the number of flows that need to be created and managed.

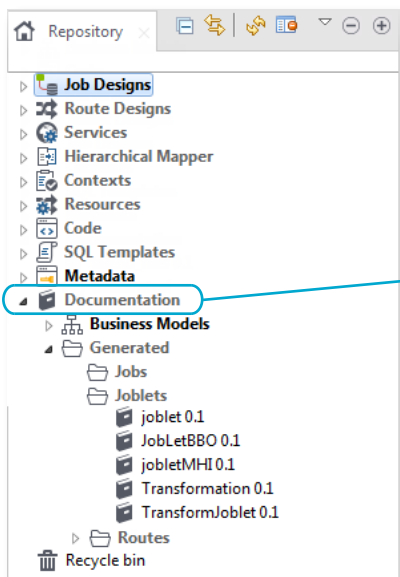


# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

**Parallelization:** To improve performance, you can partition the input data flow of a subjob into parallel processes on the same server to execute simultaneously.



**Documentation:** Often a lower priority, but critical to supporting your project, you can have project documentation auto-generated each time the job is created, updated or saved. What a times savings!



JobUSState\_0.1.html

### Component List

Component Name	Component Type
tFileInputDelimited_1	tFileInputDelimited
tLogRow_1	tLogRow

### Components Description

Component tFileInputDelimited

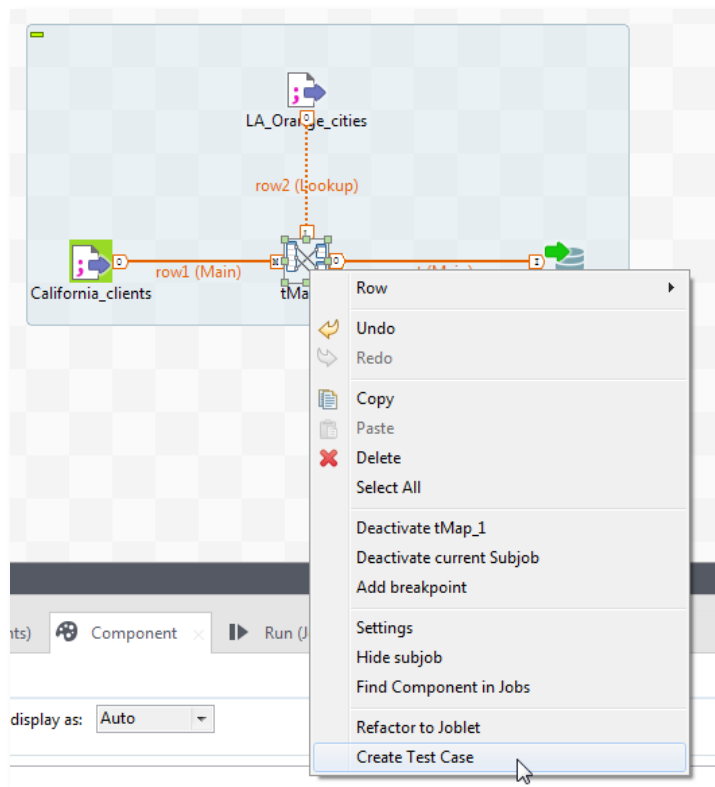
UNIQUE NAME	Label	INPUT(S)	OUTPUT(S)
tFileInputDelimited_1		none	tLogRow_1

### Component Parameters:

Properties	Values
Unique Name	tFileInputDelimited_1
Component Name	tFileInputDelimited

# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

**Creating Test Cases:** It is easy to create integration flow test cases, just click on the component and select “Create Test Case”

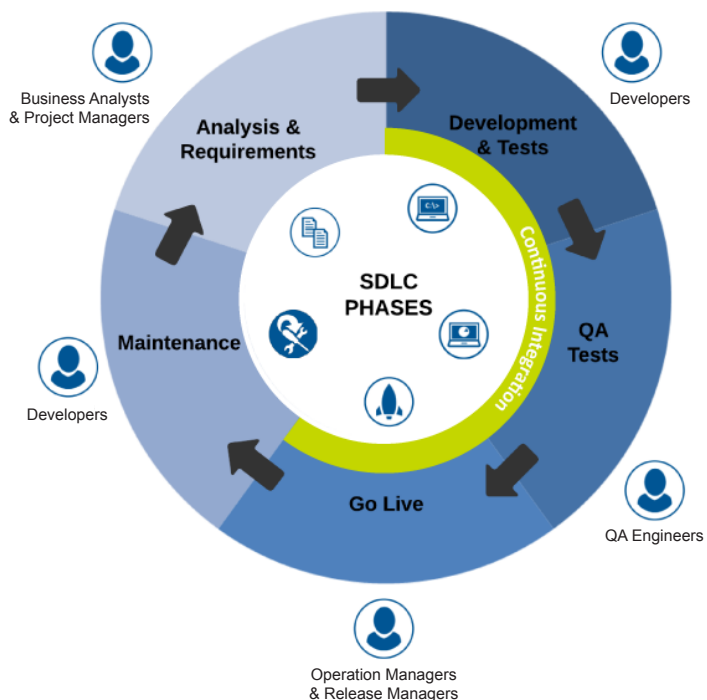


**Continuous Integration:** Get a significant DevOps productivity boost by automating build, test and deployment processes using Talend’s built-in continuous integration (CI) features.

Team members integrate work frequently using Apache Maven-based builds that incorporate a distributed revision control system (GitHub), so everyone is working on the latest updates. Git support for graphical component-level diff and merge speeds updates, while automated tests (JUnit) are generated to detect software defects earlier.

Higher quality code and fewer defects means faster time to production.

[Learn more about continuous integration >](#)  
*Video Documentation*



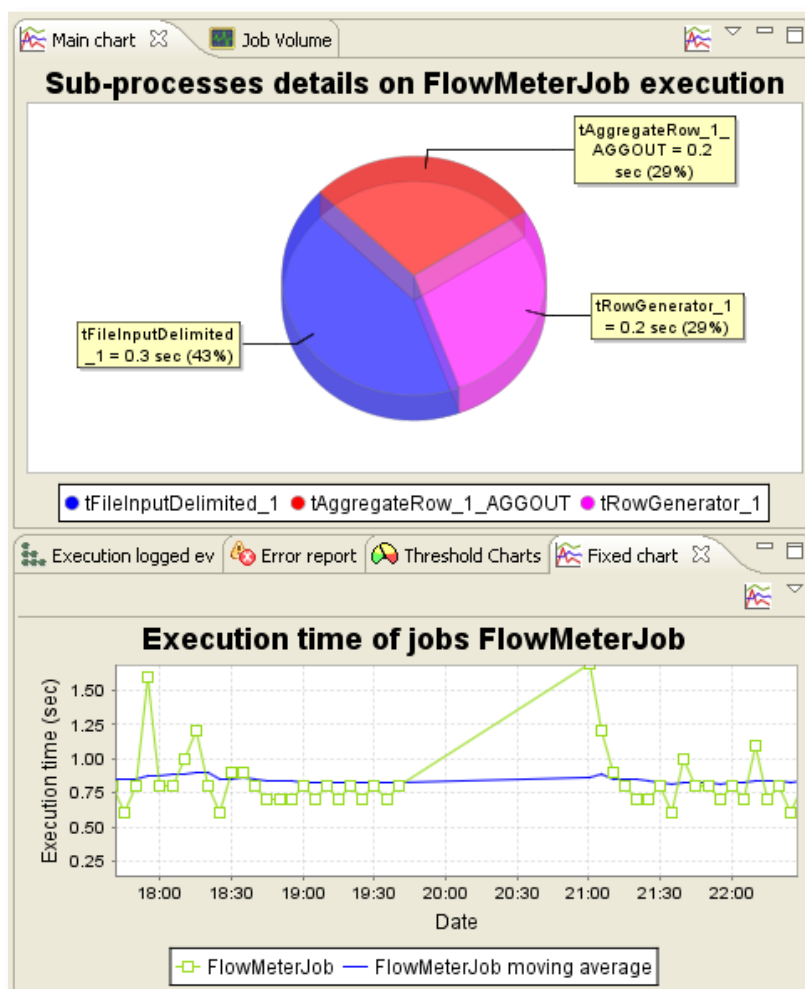
## 2. Job Monitoring and Event Logging

If your job is not performing well ... do you know why?

With Talend's Activity Monitoring Console, event logging, and Kibana web-based job monitoring dashboards, you can see detailed information of all your processes, down to the component level, to help you quickly identify bottlenecks and optimize jobs for best performance. Debug the execution of integration actions, collect exception data and transfer it to the Studio console or web dashboard.

### 2.1 Activity Monitoring Console

The Talend Activity Monitoring Console monitors your end-to-end job flow including events (successes, failures, warnings, etc.), execution times and data volumes. Through detailed monitoring you better understand the underlying component and job interaction, and can then optimize performance. For example, you can visual how long a component took to execute today, and compare execution time over the past week. [Learn more](#) >



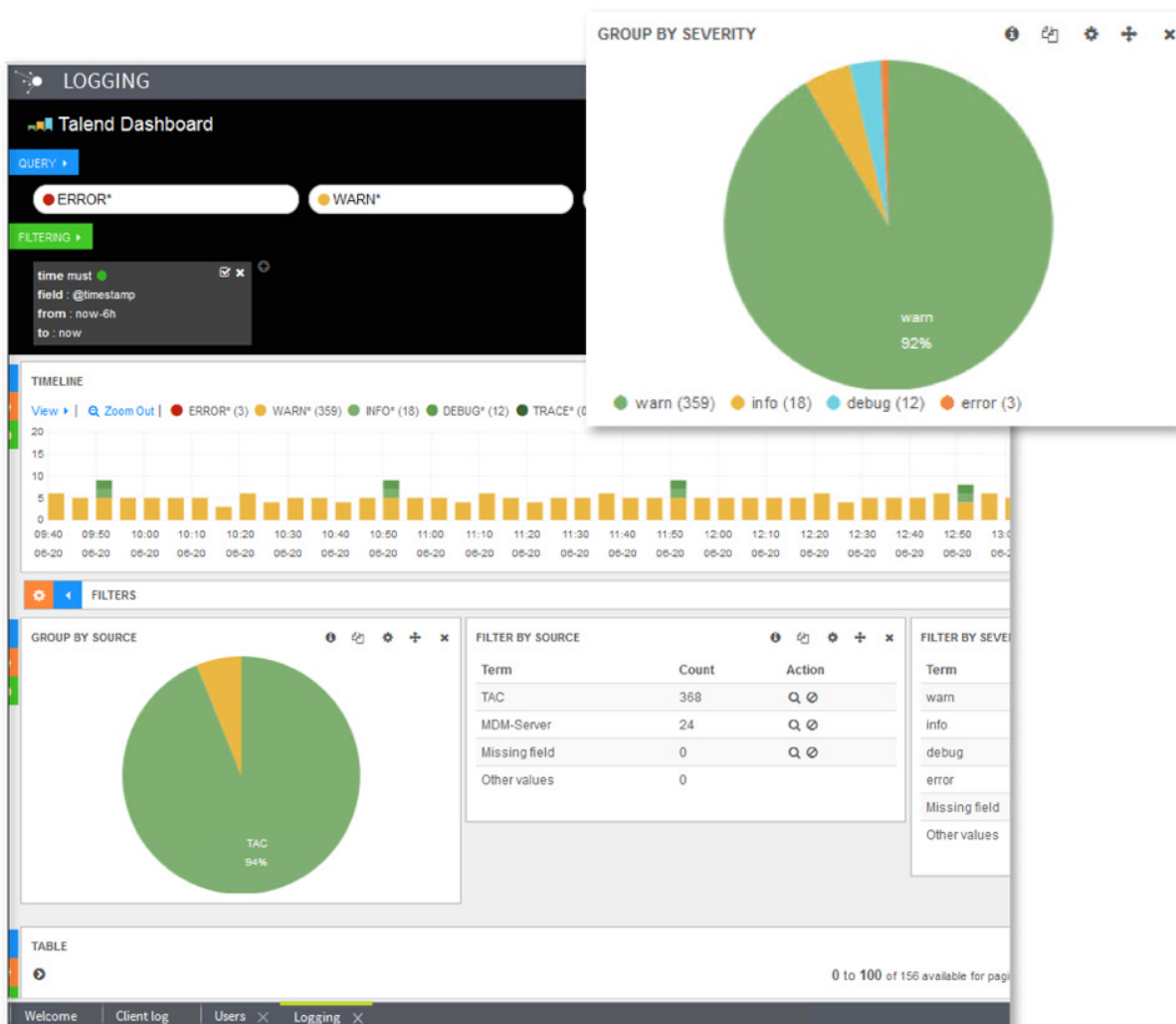
Analyze component performance in Activity Monitoring Console



# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

## 2.2 Event Logging

The Activity Monitoring Console gives you the high-level process view, and Talend event logging tools give you the details. Rather than reading thousands of lines in a log file, integrated event logging based on ELK (Elasticsearch, Logstash and Kibana) provides powerful time-series data visualization with support for custom processing (custom filtering, data enrichment, aggregation). And with Apache Log4J services for each Talend component, you can get further enhanced diagnostics and error reporting on your dashboard.

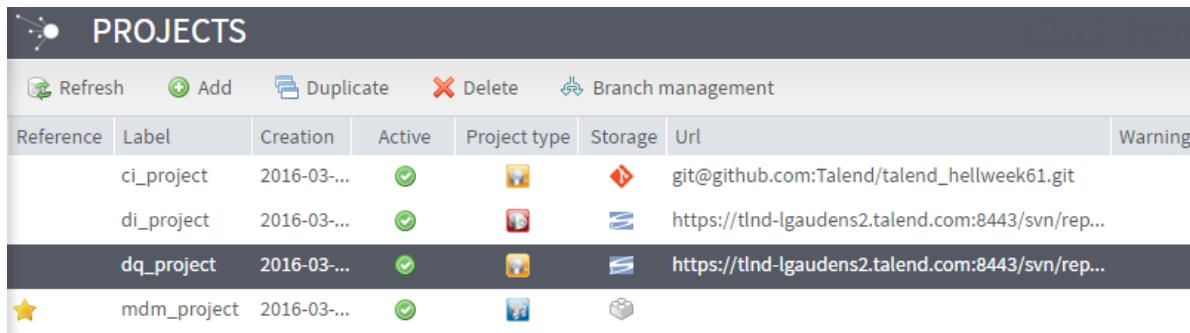


Use the default dashboards, or create customizable charts and graphs (e.g. pie chart, histogram, table) for a complete view of overall platform health, and then set up alerts so you can address issues before they happen. For example, you can build custom dashboards to monitor a unique product ID or database time-outs. [Learn more](#) >

# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

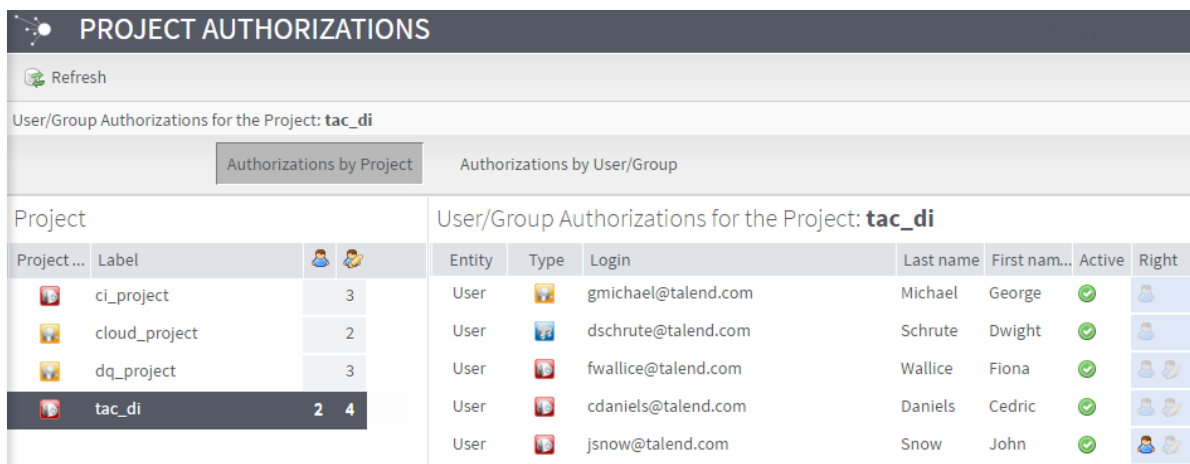
## 3. User and Deployment Management

Save significant time managing users and deployments with the Talend Administration Center, a web application where you can easily administer projects, users, access rights and licenses.



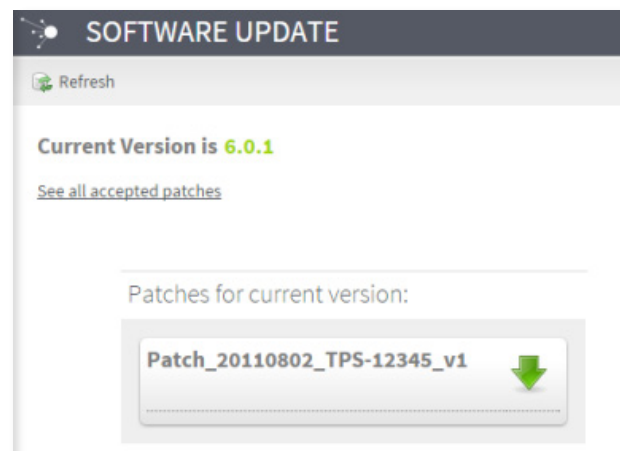
Reference	Label	Creation	Active	Project type	Storage	Url	Warning
	ci_project	2016-03-...	✓			git@github.com:Talend/talend_hellweek61.git	
	di_project	2016-03-...	✓			https://tlnd-lgaudens2.talend.com:8443/svn/rep...	
	dq_project	2016-03-...	✓			https://tlnd-lgaudens2.talend.com:8443/svn/rep...	
★	mdm_project	2016-03-...	✓				

Set authorizations on users, projects and features - you have complete control of what team members can and cannot do.



Project	Label	Entity	Type	Login	Last name	First name	Active	Right
ci_project		User		gmichael@talend.com	Michael	George	✓	
cloud_project		User		dschrute@talend.com	Schrute	Dwight	✓	
dq_project		User		fwallice@talend.com	Wallice	Fiona	✓	
tac_di		User		cdaniels@talend.com	Daniels	Cedric	✓	
		User		jsnow@talend.com	Snow	John	✓	

Checking for and getting software updates is simplified, as it is done straight through the tool.

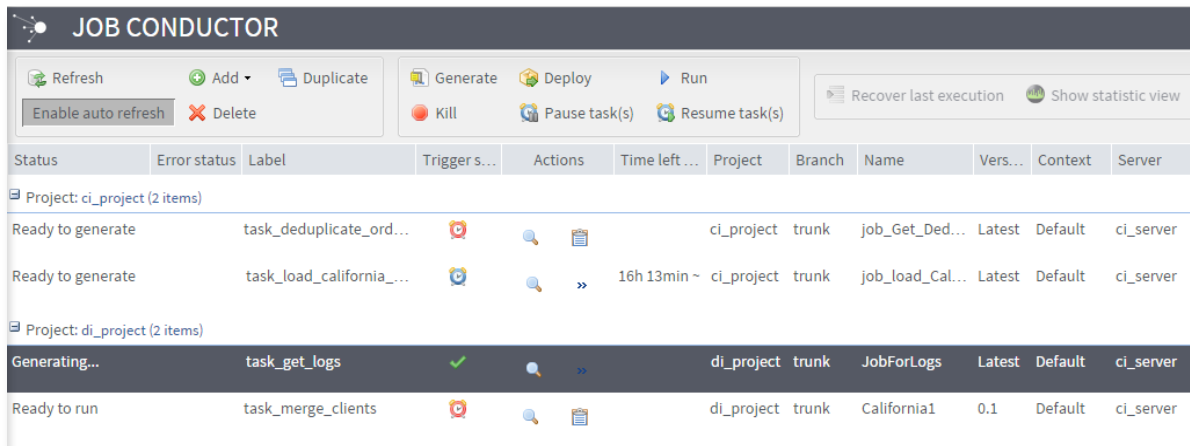


Current Version is 6.0.1
See all accepted patches
Patches for current version:
Patch_20110802_TPS-12345_v1



# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

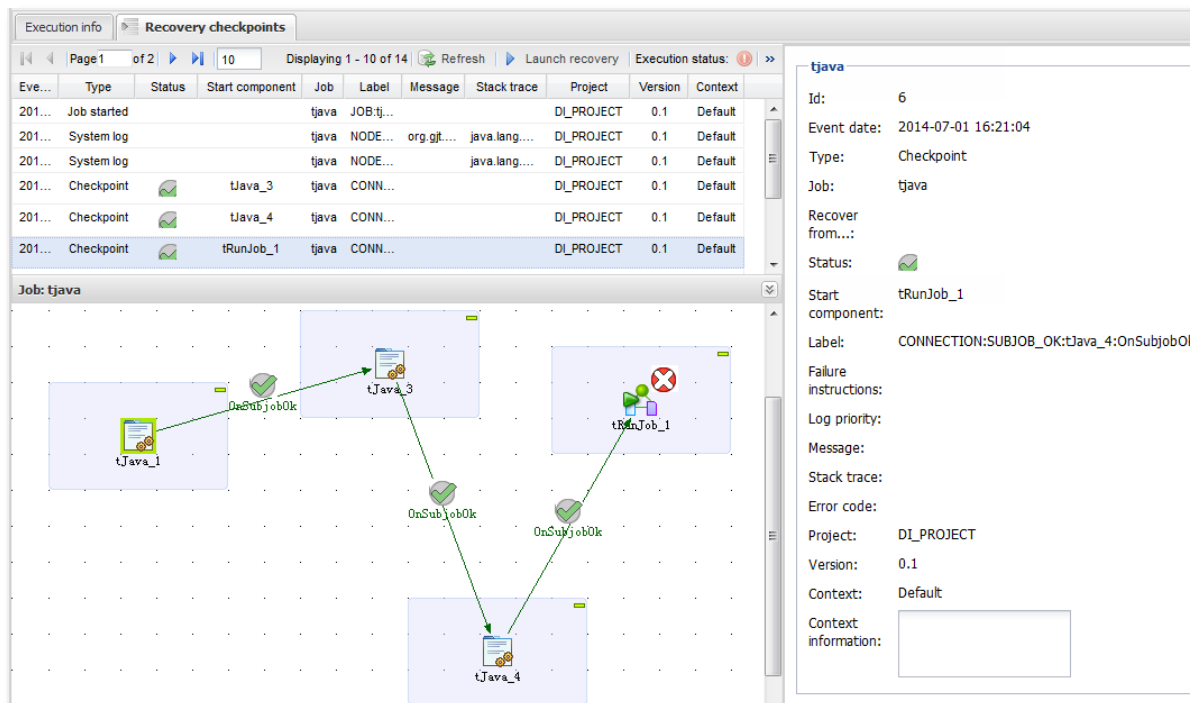
**Job Scheduling:** For deployment management, execution (job) servers are configured and managed in the graphical console. Next, graphically schedule and manage integration jobs with the Job Conductor, where you can define servers to execute on, time to start and stop, triggers for task execution order, resource (memory, process allocation) management, and more. Job scheduling can be parameterized with context variables or use the API for programmatic control.



Status	Error status	Label	Trigger s...	Actions	Time left ...	Project	Branch	Name	Vers...	Context	Server
Project: ci_project (2 items)											
Ready to generate		task_deduplicate_ord...				ci_project	trunk	job_Get_Ded...	Latest	Default	ci_server
Ready to generate		task_load_california_...			16h 13min ~	ci_project	trunk	job_load_Cal...	Latest	Default	ci_server
Project: di_project (2 items)											
Generating...		task_get_logs				di_project	trunk	JobForLogs	Latest	Default	ci_server
Ready to run		task_merge_clients				di_project	trunk	California1	0.1	Default	ci_server

*Manage integration deployments with Job Conductor*

**Job Restart:** If a job should fail, you can easily restart execution from the latest checkpoint before the failure.



Execution info Recovery checkpoints

Page 1 of 2 10 Displaying 1 - 10 of 14 Refresh Launch recovery Execution status: 1

Event	Type	Status	Start component	Job	Label	Message	Stack trace	Project	Version	Context
201...	Job started			tjava	JOB:tj...			DI_PROJECT	0.1	Default
201...	System log			tjava	NODE...	org.gjt...	java.lang...	DI_PROJECT	0.1	Default
201...	System log			tjava	NODE...		java.lang...	DI_PROJECT	0.1	Default
201...	Checkpoint		tJava_3	tjava	CONN...			DI_PROJECT	0.1	Default
201...	Checkpoint		tJava_4	tjava	CONN...			DI_PROJECT	0.1	Default
201...	Checkpoint		tRunJob_1	tjava	CONN...			DI_PROJECT	0.1	Default

Job: tjava

Diagram showing job flow: tJava\_1 -> tJava\_3 -> tJava\_4 -> tRunJob\_1. Checkpoints are marked with green checkmarks.

Job details:

- Id: 6
- Event date: 2014-07-01 16:21:04
- Type: Checkpoint
- Job: tjava
- Recover from...:
- Status:
- Start component: tRunJob\_1
- Label: CONNECTION:SUBJOB\_OK:tJava\_4:OnSubjobOk
- Failure instructions:
- Log priority:
- Message:
- Stack trace:
- Error code:
- Project: DI\_PROJECT
- Version: 0.1
- Context: Default
- Context information:

[Learn more about Talend Administration Center](#) > [Video Documentation](#)

# Just the Facts: Technical Differences Between Talend Data Integration and Talend Open Studio

## 4. Integrated Data Preparation and Data Stewardship

Developers can save hours by allowing business users to access data and do basic cleansing and enrichment. End users know their data best, so let them do the work through an easy-to-use web interface, while IT governs the use of data and facilitates collaboration across the enterprise.

Talend Data Preparation enables anyone to access and cleanse data using browser-based, point-and-click tools. Smart guides and visual tools help anyone quickly understand data attributes and quality status. Results can easily be exported into desktop tools such as Tableau and Excel, enabling immediate decision-making. With one click you can then embed data preparation recipes into data integration pipelines.

Talend Data Stewardship provides robust data curation tools to help you quickly identify, manage, and resolve data integrity issues. Teams work together to establish a single version of the truth.

The diagram illustrates the roles of Data Analysts, Business, and Stewards in using Talend Data Integration. The interface shows a table of employee data with filters and an export menu.

**Roles:**

- Data Analysts:** Represented by a group of icons and a central profile picture.
- Business:** Represented by a group of icons and a central profile picture.
- Stewards:** Represented by a group of icons and a central profile picture.

**Interface Elements:**

- Header: "Your export is done. Click [here](#) to proceed."
- Filters: "New filter...", "job\_title = chemi", "date in Feb 2015".
- Table Columns: #, ID, FIRST\_NAME, LAST\_NAME, EMAIL, JOB\_TITLE, COMPANY, CITY.
- Table Rows: 309, 505, 699, 1309, 1505, 1699, 2309, 2505, 2699, 3309, 3505.
- Export Menu: "Export", "Last XLSX Export (5 minutes ago)", "Last CSV Export (5 minutes ago)".

**LIVE DATA-SET**

**Integrations:** ORACLE, XLSX, salesforce, TERADATA, Marketo, amazon web services™, ...and more

[Learn more](#) > Video Documentation



## 5. And More ....

The appendix on the following page highlights all the differences between Talend Open Studio and Talend Data Integration. Depending on your requirements, there may be other Talend products that fit your needs, and they all leverage the same Talend Studio you are familiar with.

- If data quality is important, then Talend Data Management Platform extends Talend Data Integration with data quality and governance features.
- If you want to reduce deployment and management costs further, Talend Integration Cloud is a secure, managed integration Platform-as-a-Service (iPaaS), where you design jobs in Talend Studio, but instead of deploying and managing on-premises, you publish as a service to a secure AWS workspace managed by Talend.
- If you need to integrate big data sources and distributions and generate optimized Spark or MapReduce code, then check out Talend Big Data



## Appendix: A Comparison of Talend Data Integration Offerings

Talend 6.4

	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
License	Apache	Subscription	Subscription
Productivity Increasing Graphical Wizards and Tools			
900+ components and connectors	✓	✓	✓
Eclipse-based developer tooling	✓	✓	✓
ETL & ELT support	✓	✓	✓
Job designer	✓	✓	✓
Versioning	✓	✓	✓
Audit		✓	✓
Auto-generation documentation		✓	✓
Change data capture (CDC)		✓	✓
Drools business rule management system		✓	✓
Distant run		✓	✓
Dynamic schema		✓	✓
Interactive data viewer		✓	✓
Metadata bridge for metadata import/export		✓	✓
Parallelization		✓	✓
Re-usable joblets		✓	✓
Advanced SAP integration		✓	✓
Testing, debugging & tuning tools		✓	✓
Wizards		✓	✓
Visual mapping for complex XML and EDI			✓

# A Comparison of Talend Data Integration Offerings



	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
<b>Improved Team Productivity</b>			
Continuous delivery data integration		✓	✓
Impact analysis		✓	✓
Jobs compare		✓	✓
Reference projects		✓	✓
Team collaboration with shared repository		✓	✓
Centralized metadata management		✓	✓
Repository manager			✓
<b>Monitor and Manage Complex Deployments with Ease</b>			
Amazon EC2 lifecycle control		✓	✓
Check points, error recovery		✓	✓
Context management (dev, QA, prod)		✓	✓
Deployment manager		✓	✓
Execution plan, time and event-based scheduler		✓	✓
Log server with dashboard		✓	✓
Activity Monitoring Console		✓	✓
Talend Administration Center		✓	✓
High availability, load balancing, failover			✓
<b>Accelerate Data Usage with Data Preparation and Stewardship</b>			
Import, export & combine data from any database, Excel or CSV file		✓	✓
Import, export & combine CSV, Parquet & AVRO files from/to HDFS		✓	✓

# A Comparison of Talend Data Integration Offerings



	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
Export to Tableau		✓	✓
Self-service on-demand access to sanctioned datasets		✓	✓
Share data preparations & datasets		✓	✓
Operationalize preparations into any data & cloud integration flow		✓	✓
Run preparations on Apache Beam		✓	✓
Auto-discovery, profiling, smart suggestions, and data visualization		✓	✓
Auto-discovery & auto-profiling of custom semantic types		✓	✓
Smart & selective sampling & full-runs		✓	✓
Data tracking & masking with role-based security		✓	✓
Cleansing and enrichment functions		✓	✓
Data Stewardship App for data curation and certification		✓	✓
<b>Data Quality and Governance</b>			
Data profiling & analytics with graphical charts & drilldown data			✓
Automate data quality error resolution and enforce rules			✓
Data masking			✓
Data quality portal with monitoring, reporting & dashboards			✓
Semantic discovery with automatic detection of patterns			✓
Comprehensive survivorship			✓
Data sampling			✓
Enrichment, harmonization, fuzzy matching & de-duplication			✓



## A Comparison of Talend Data Integration Offerings

	Open Studio for Data Integration	Talend Data Integration	Talend Data Management Platform
<b>Advanced Data Profiling</b>			
Fraud pattern detection using Benford Law			✓
Advanced statistics with indicator thresholds			✓
Column set analysis			✓
Advanced matching analysis			✓
Time column correlation analysis			✓
<b>Reduce Risk with Product Indemnification</b>		✓	✓